COVID-CT-Dataset: A CT Scan Dataset about COVID-19

Abstract

During the outbreak time of COVID-19, computed tomography (CT) is a useful manner for diagnosing COVID-19 patients. Due to privacy issues, publicly available COVID-19 CT datasets are highly difficult to obtain, which hinders the research and development of AI-powered diagnosis methods of COVID-19 based on CTs. To address this issue, we build an open-sourced dataset -- COVID-CT, which contains 349 COVID-19 CT images from 216 patients and 463 non-COVID-19 CTs. The utility of this dataset is confirmed by a senior radiologist who has been diagnosing and treating COVID-19 patients since the outbreak of this pandemic. We also perform experimental studies which further demonstrate that this dataset is useful for developing AI-based diagnosis models of COVID-19. Using this dataset, we develop diagnosis methods based on multi-task learning and self-supervised learning, that achieve an F1 of 0.90, an AUC of 0.98, and an accuracy of 0.89. According to the senior radiologist, models with such performance are good enough for clinical usage. The data and code are available at this https URL